ELSEVIER
Theoretical Computer Science 287 (2002) 393–418
Theoretical Computer Science
www.elsevier.com/locate/tcs

On list update and work function algorithms

Eric J. Anderson[a,*], Kirsten Hildrum[b], Anna R. Karlin[a], April Rasala[c], Michael Saks[d]

[a] *Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195–2350, USA*
[b] *Computer Science Division, University of California, Berkeley, USA*
[c] *Laboratory of Computer Science, Massachusetts Institute of Technology, USA*
[d] *Department of Mathematics, Rutgers University, USA*

Received March 2001; accepted May 2001

Abstract

The *list update* problem, a well-studied problem in dynamic data structures, can be described abstractly as a metrical task system. In this paper, we prove that a generic metrical task system algorithm, called the *work function algorithm*, has constant competitive ratio for list update. In the process, we present a new formulation of the well-known "list factoring" technique in terms of a partial order on the elements of the list. This approach leads to a new simple proof that a large class of online algorithms, including Move-To-Front, is $(2 - 1/k)$-competitive, for k the list length.

1. Introduction

1.1. Motivation

The *list accessing* or *list update* problem is one of the most well-studied problems in competitive analysis [1, 4, 5, 8, 11]. The problem consists of maintaining a set S of items in an unsorted linked list, for example as a data structure for implementation of a dictionary. The data structure must support three types of requests: ACCESS(x), INSERT(x) and DELETE(x), where x is the name, or "key", of an item stored in the list. We associate a cost with each of these operations as follows: accessing or deleting the ith item on the list costs i; inserting a new item costs $j + 1$ where j is the number of items currently on the list before insertion. We also allow the list to be

* Corresponding author.
E-mail addresses: eric@cs.washington.edu (E.J. Anderson), hildrum@cs.berkeley.edu (K. Hildrum), karlin@cs.washington.edu (A.R. Karlin), arasala@theory.lcs.mit.edu (A. Rasala), saks@math.rutgers.edu (M. Saks).

PII: S0304-3975(01)00253-5

reorganized, at a cost measured in terms of the minimum number of transpositions of consecutive items needed for the reorganization. We consider the standard cost model in the literature: immediately after an access or an insertion, the requested item may be moved at no extra cost to a position closer to the front of the list. These exchanges are called *free exchanges*. Intuitively, using free exchanges, the algorithm can lower the cost on subsequent requests. In addition, at any time, two adjacent items in the list can be exchanged at a cost of 1. These exchanges are called *paid exchanges*. The list update problem is to devise an algorithm for reorganizing the list, by performing free and/or paid exchanges, that minimizes search and reorganization costs. As usual, the algorithm will be evaluated in terms of its competitive ratio.

As with much of the work on list accessing, we will focus on the *static list update problem*, where the list starts out with some number k elements in it, and all requests are accesses. The results described are easily extended to the dynamic case including insertions and deletions. Specifically, the cost of an insertion is the same for any algorithm; and the cost of a deletion is the same as the cost of an access. Some results for static list update are expressed in terms of the length k of the list. In the case of dynamic list update, the length k is no longer uniquely defined. However, for constant-competitive ratio results, it is enough for our purposes to interpret k as the *maximum* length of the list where necessary.

Many deterministic online algorithms have been proposed for the list update problem. Of these, perhaps the most well known is the *Move-To-Front* algorithm: after accessing an item, move it to the front of the list, without changing the relative order of the other items. *Move-To-Front* is known to be $2-2/(k+1)$ competitive, and this is best possible [8, 12].

We note that other cost models have also been considered for the list update problem [10, 15, 16]. Increasing the cost of exchanges to two (instead of one) makes *Move-To-Front* optimal; this provides an independent proof that *Move-To-Front* is two-competitive. Other alternatives analyzed in the literature include allowing free exchanges for other than the referenced element, and allowing free exchanges between elements that are not adjacent [9, 10]. These alternative cost models can lead to qualitatively different results.

1.2. Metrical task systems

The (static) list update problem can also be considered within the *metrical task system* framework introduced by Borodin et al. [6]. Metrical task systems (MTS) are an abstract model for online computation that captures a wide variety of online problems (paging, list update and the k-server problem, to name a few) as special cases. A metrical task system is a system with n states, with a distance function d defined on the states: $d(i,j)$ is the distance between states i and j. The distances are assumed to form a metric. The MTS has a set $\mathscr{T}$ of allowable tasks; with each task $\tau \in \mathscr{T}$ is associated a vector $(\tau(1), \tau(2), \dots, \tau(n))$ where $\tau(i)$ is the (nonnegative) cost of processing task τ in state i. An online algorithm is given a starting state and a sequence $\sigma = \sigma_1, \ldots, \sigma_n$ of tasks to be processed online, and must decide in which state to process each task. The goal of the algorithm is to minimize the total distance moved plus the total processing costs. The cost of the online algorithm is compared to that of an optimal algorithm, which produces an optimal sequence $s_0, s_1, \ldots, s_n$ of states, one for which the cost $d(s_0, s_1) + \sigma_1(s_1) + \cdots + \sigma_n(s_n)$ is minimized.

The list update problem can be viewed as a metrical task system as follows. The states of the list update MTS are the $k!$ possible orderings of the k elements in the list, which we also call *list configurations*. There are k possible tasks, one corresponding to each list element x. The cost $\tau_x(\pi)$ of processing the task τ_x in a particular list configuration π, is equal to the depth of x in the list π. The distance between two states or list configurations is the number of inversions between the list orderings, considered as permutations. In this formulation, an algorithm produces a sequence of pairwise inversions of adjacent elements in the list, punctuated by a sequence of reference points at which the references $\sigma = \sigma_1, \ldots,$ are made. The cost of such a sequence is the number of the pairwise inversions, plus the total cost of each of the references at the corresponding reference points. An optimal sequence is a sequence of inversions and reference points that minimizes the total cost. [1]

The traditional description of list update in terms of "free exchanges" and "paid exchanges" is identical in cost to this model. See [15, Theorem 1]. Any sequence containing "free exchanges" can be translated to the MTS model by treating such exchanges as instead made at unit cost immediately before the item is referenced. (This translation consists of simply moving the reference point to immediately after the "free exchanges".) The cost of these exchanges is precisely offset by a lower reference cost for the referenced element. We continue to use the terminology of "free exchanges" to describe those exchanges involving the next-referenced element in the MTS model, and "paid exchanges" to describe those exchanges not involving the next-referenced element.

One of the initial results about metrical task systems was that the *work function algorithm* (*WFA*) has competitive ratio $2n-1$ for all MTSs, where n is the number of states in the metrical task system [6]. It was also shown that this is best possible, in the sense that there exist metrical task systems for which no online algorithm can achieve a competitive ratio lower than $2n - 1$. However, for many MTSs the upper bound of $2n - 1$ is significantly higher than the best achievable competitive ratio. For example, there are known constant-competitive algorithms for list update, even though the MTS for a list of k elements has $k!$ states. Another example is the k-server problem on a finite metric space consisting of r points. For this problem, the metrical task system has $n = \binom{r}{k}$ states, but a celebrated result of Koutsoupias and Papadimitriou shows that in fact the *very same work function algorithm* is $2k - 1$ competitive for this problem [13], nearly matching the known lower bound of k on the competitive ratio [14].

[1] Hereafter we use the terms "exchanges", "inversions", "transpositions", and "interchanges" interchangeably to mean specifically the transposition of adjacent elements in the list.

Unfortunately, our community understands very little at this point about how to design competitive algorithms that achieve close to the best possible competitive ratio for broad classes of metrical task systems. Indeed, one of the most intriguing open questions in this area is

> For what metrical task systems is the work function algorithm strongly competitive? [2]

Burley and Irani have shown the existence of metrical task systems for which the work function algorithm is not strongly competitive [8]. However, these "bad" metrical task systems seem to be rather contrived, and it is widely believed that the work function algorithm is in fact strongly competitive for large classes of natural metrical task systems. The desire to make progress towards answering this broad question is the foremost motivation for the work described in this paper. We were specifically led to reconsider the list update problem when we observed the following curious fact:

> The *Move-To-Front* algorithm for list update is a work function algorithm. (Proposition 8, Section 4.)

This observation was intriguing for two reasons. First because it raised the question of whether work function algorithms generally (that is, those with more general tie-breaking rules than that used in *Move-To-Front*) are strongly competitive for list update. This would provide an example of a substantially different type of metrical task system for which the work function algorithm is strongly competitive than those considered in the past.

The second and perhaps more exciting reason for studying work functions as they relate to list update is the tantalizing possibility that insight gained from that study could be helpful in the study of dynamic optimality for self-adjusting binary search trees [4, 19]. It is a long-standing open question whether or not there is a strongly competitive algorithm for dynamically rearranging a binary search tree using rotations, in response to a sequence of accesses. The similarity between *Move-To-Front* as an algorithm for dynamically rearranging linked lists, and the splay tree algorithm of Sleator and Tarjan [19] for dynamically rearranging binary search trees, long conjectured to be strongly competitive, is appealing. Our hope is that the use of work function-like algorithms might help to resolve this question for self-adjusting binary search trees.

1.3. Results

The main result of this paper is a proof that a class of work function algorithms is O(1)-competitive for the list update problem. [3] Proving this theorem requires getting a handle on the work function values, the optimal offline costs of ending up in each state. This is tricky, as the offline problem is very poorly understood. At present it

[2] We say an algorithm is *strongly competitive* if its competitive ratio is within a constant factor of the best competitive ratio achievable.

[3] The proof does not achieve the best possible competitive ratio of 2.

is even unknown whether the problem of computing the optimal cost of executing a request sequence is NP-hard. The fastest optimal off-line algorithm currently known runs in time $O(2^k k!n)$, where k is the length of the list and $n=|\sigma|$ is the length of the request sequence σ [15].

Using the framework that we have developed for studying work functions and list update, we also present a new simple and illustrative proof that *Move-To-Front* and a large class of other online algorithms are $(2-1/k)$-competitive.

The rest of the paper is organized as follows. In Section 2, we present background material on work functions and on the work function algorithm. In Section 3, we present a formulation of the list update work functions in terms of a partial order on the elements of the list and use this formulation to prove that a large class of list update algorithms are $(2-1/k)$-competitive. Finally, in Section 4 we present our main result, that work function algorithms are strongly competitive for list update.

2. Background

We begin with background material on work functions and work function algorithms. As customary, we focus on the *competitive ratio* of the online algorithms. An algorithm A has competitive ratio c if for all sequences σ its cost $A(\sigma)$ is bounded by $A(\sigma)\leqslant cOPT(\sigma)+K$ for some additive constant K.

2.1. Work functions

Consider an arbitrary metrical task system, with states $s\in S$ and tasks $\tau\in T$. Given a sequence of requests σ, denote the tth request in the sequence as σ_t. Let the task σ_{t+1} be denoted by τ. Let $\tau(s)$ denote the cost of executing task τ in state s.

Definition 1. The *work function* $\omega_t(s)$ for any state s and index t is the lowest cost of satisfying the first t requests of σ and ending up in state s [6, 9].

Because the states and task costs are time-independent, the work functions can be calculated through a dynamic programming formulation (which can equally be taken as the definition):

$$\omega_{t+1}(s') = \min_s (\omega_t(s) + \tau(s) + d(s,s')). \tag{1}$$

This equation can be interpreted by observing that any optimal sequence of states $s_0, s_1, \ldots, s = s_{t+1}, s'$ achieving $\omega_{t+1}(s')$ must have satisfied $\tau=\sigma_{t+1}$ in some state s, incurring $\omega_t(s)$ up to that point, and $\tau(s)+d(s,s')$ thereafter. Because any optimal sequence achieving $\omega_t(s)$ can be substituted, the dynamic programming formulation is appropriate. In particular, for any s', the state s satisfies $\omega_{t+1}(s)=\omega_t(s)+\tau(s)$. We identify this property of *fundamental* states, which will be convenient in later defining the work function algorithm:

Definition 2. A state f is *fundamental at time t* if it satisfies $\omega_{t+1}(f)=\omega_t(f)+\tau(f)$.

(Where the context is evident, we will simply say a state f is "fundamental".)

We next note several elementary identities, which hold generally for metrical task systems, at all times t and for all states s and s'. As above, we let τ denote the $(t+1)$st task σ_{t+1}, and $\tau(s)$ its task cost in the state s.

Proposition 1.

$$\omega_t(s) \leqslant \omega_t(s') + d(s,s').$$

Proof. For notational convenience, we show $\omega_{t+1}(s) \leqslant \omega_{t+1}(s') + d(s,s')$. From the definition (Eq. (1)), there is some $\tilde{s}$ for which $\omega_{t+1}(s') = \omega_t(\tilde{s}) + \tau(\tilde{s}) + d(\tilde{s},s')$. We know that $\omega_t(\tilde{s}) + \tau(\tilde{s}) + d(\tilde{s},s)$ is an upper bound on $\omega_{t+1}(s)$ (by Eq. (1)). By the triangle inequality on the metric d, $d(\tilde{s},s) \leqslant d(\tilde{s},s') + d(s',s)$. So we have $\omega_{t+1}(s) \leqslant \omega_{t+1}(s') + d(s,s')$. $\square$

Proposition 2.

$$\omega_{t+1}(s) \geqslant \omega_t(s).$$

Proof. By the alternative definition above (Eq. (1)), for some s' we have $\omega_{t+1}(s) = \omega(s') + d(s,s') + \tau(s')$. By Proposition 1, $\omega_t(s) \leqslant \omega(s') + d(s,s')$. Since all task costs are nonnegative, $\tau(s') \geqslant 0$, and the result follows. $\square$

Proposition 3.

$$\omega_{t+1}(s) \leqslant \omega_t(s) + \tau(s).$$

Proof. By the definition (Eq. (1)), $\omega_{t+1}(s) = \min_{s'}(\omega_t(s') + \tau(s') + d(s,s'))$, so for all such s', $\omega_{t+1}(s) \leqslant \omega_t(s') + \tau(s') + d(s,s')$. Substituting s for s', and noting that $d(s,s) = 0$, the result follows. $\square$

Proposition 4. *For any* s,

$$\omega_{t+1}(s) = \omega_{t+1}(f) + d(f,s)$$

for some state f *that is fundamental at time* t. (*The state* s *is derived from some fundamental state.*)

Proof. By the definition (Eq. (1)), there is some f for which $\omega_{t+1}(s) = \omega_t(f) + \tau(f) + d(f,s)$. We want to show that this f is fundamental. By Proposition 1, $\omega_{t+1}(s) \leqslant \omega_{t+1}(f) + d(f,s)$, so $\omega_t(f) + \tau(f) \leqslant \omega_{t+1}(f)$. But $\omega_t(f) + \tau(f) \geqslant \omega_{t+1}(f)$ by Proposition 3. Hence $\omega_t(f) + \tau(f) = \omega_{t+1}(f)$ and f is fundamental at time t. Then $\omega_{t+1}(s) = \omega_{t+1}(f) + d(f,s)$ by substitution. $\square$

2.2. The work function algorithm

The *WFA* [6, 9], defined for an arbitrary metrical task system, is the following:

Definition 3. *WFA*: When in state s_t, service the request $\sigma_{t+1} = \tau$ in the state s_{t+1} such that

$$s_{t+1} = \operatorname*{argmin}_s (\omega_{t+1}(s) + d(s_t, s)),$$

where the minimum is taken over states s that are fundamental at time t.

From Definition 2, we see that the work function algorithm chooses s_{t+1} so that

$$s_{t+1} = \operatorname*{argmin}_s (\omega_t(s) + \tau(s) + d(s_t, s)). \tag{2}$$

We consider a variant of this work function algorithm, differing only in the subscript of the work function:

Definition 4. *WFA′*: When in state s_t, service the request τ in the state s_{t+1} such that

$$s_{t+1} = \operatorname*{argmin}_s (\omega_{t+1}(s) + \tau(s) + d(s_t, s)).$$

In this definition, the state s need not be fundamental at time t. (See Proposition 5 below.)[4]

The minimum in this expression may not be unique. Accordingly, we define the class of states to which the work function algorithm might move.

Definition 5. Given that *WFA′* visits state s_t at time t, a state s at time $t+1$ is *wfa-eligible* if it is one of the states that minimizes the expression in Definition 4.

We establish the following properties for wfa-eligible states.

Proposition 5. *Suppose WFA′ is in state s_t at time t. Suppose s is wfa-eligible at time t, and suppose further that $\omega_{t+1}(s) = \omega_{t+1}(f) + d(f, s)$ where f is fundamental.* (*There is at least one such state f by Proposition* 4.) *Then f is also wfa-eligible at time t, and $\tau(f) = \tau(s)$.* (*The fundamental state f is wfa-eligible if s is.*)

Proof. Since s is wfa-eligible, it minimizes the expression in Definition 4, $\omega_{t+1}(s) + \tau(s) + d(s_t, s) \leqslant \omega_{t+1}(s') + \tau(s') + d(s_t, s')$ for all s'. If we show that $\omega_{t+1}(f) + \tau(f) + d(s_t, f) \leqslant \omega_{t+1}(s) + \tau(s) + d(s_t, s)$, then f minimizes that expression as well, and f then must also be wfa-eligible.

We observe first that $\tau(f) \leqslant \tau(s)$. By Propositions 2 and 3, $\omega_{t+1}(s) \leqslant \omega_t(s) + \tau(s) \leqslant \omega_t(f) + \tau(s) + d(f, s)$. Then $\tau(s) < \tau(f)$ would imply $\omega_{t+1}(s) < \omega_t(f) + \tau(f) + d(f, s) = \omega_{t+1}(f) + d(f, s)$. By hypothesis, however, we have $\omega_{t+1}(f) + d(f, s) = \omega_{t+1}(s)$.

Next, by the triangle inequality, $d(s_t, f) \leqslant d(s_t, s) + d(f, s)$. Then $\omega_{t+1}(f) + d(s_t, f) \leqslant \omega_{t+1}(f) + d(s_t, s) + d(f, s) = \omega_{t+1}(s) + d(s_t, s)$. Since $\tau(f) \leqslant \tau(s)$, we have $\omega_{t+1}(f) + d(s_t, f) + \tau(f) \leqslant \omega_{t+1}(s) + d(s_t, s) + \tau(s)$, and f is wfa-eligible.

[4] If the state s is fundamental, then the quantity minimized is equal to $\omega_t(s) + 2\tau(s) + d(s_t, s)$, but we do not make any special use of this fact.

Finally, since s is also wfa-eligible, the above inequality cannot be strict. It would be if $\tau(f) < \tau(s)$, so we must have $\tau(f) = \tau(s)$. □

Proposition 6. *If s is wfa-eligible, then $\tau(s) \leqslant \tau(s_t)$.*

Proof. Suppose instead that $\tau(s) > \tau(s_t)$. Then the condition for s to be wfa-eligible (Definition 4) is $\omega_{t+1}(s) + \tau(s) + d(s, s_t) > \omega_{t+1}(s) + \tau(s_t) + d(s, s_t) \geqslant \omega_{t+1}(s_t) + \tau(s_t)$ by Proposition 1. But this last expression is Definition 4 applied to the state s_t. If s_t satisfies Definition 4 strictly more strongly than s, s cannot be wfa-eligible. □

We will see that, when applying WFA' to list update, there always exists at least one wfa-eligible state that requires no paid exchanges (Proposition 8). In the remainder of the paper, we will assume that WFA' chooses to move to a wfa-eligible state of this type, i.e., one that can be reached by moving the referenced element only. That is, in what follows we consider only work function algorithms that perform “free exchanges”.

2.3. Observations

The work function algorithm can be viewed as a compromise between two very natural algorithms. First, a natural *greedy* algorithm tries to minimize the cost spent on the current step. It services the $(t+1)$st request τ in a state s that minimizes $d(s_t, s) + \tau(s)$. Another natural algorithm is a *retrospective* algorithm, which tries to match the state chosen by the optimal offline algorithm. It services the $(t+1)$st request τ in a state s that minimizes $\omega_{t+1}(s)$.

Each of these natural algorithms is known to be noncompetitive for many metrical task systems. WFA combines these approaches and, interestingly, this results in an algorithm which is known to be strongly competitive for a number of problems for which neither the *greedy* and *retrospective* algorithms are competitive. [5]

The difference between WFA and the variant, WFA', is in the subscript of the work function. We actually feel that WFA' is a slightly more natural algorithm, in light of the discussion above about combining a greedy approach and a retrospective approach. It is this latter work function algorithm WFA' that we will focus on in this paper. It is fairly easy to extend our proof that WFA' is O(1)-competitive for list update to handle WFA as well. [6]

[5] Varying the relative weighting of the greedy and retrospective components of the work function algorithm was explored in [7].

[6] In addition, many prior results which hold for WFA also hold for WFA'. For example, for the k-server problem the work function values at t and $t+1$ are identical for any states s that serve the $t+1$st request, $\omega_{t+1}(s) = \omega_t(s)$. Hence WFA' and WFA define the same algorithm, and so WFA' is $2k-1$ competitive for the k-server problem. The proof that WFA is $2n-1$ competitive for any metrical task system with n states also holds for WFA' (using the same potential function), and so WFA' also is $2n-1$ competitive for any metrical task system.

3. A different view on list factoring

A technique which has been used in the past to analyze list update algorithms is the *list factoring* technique, which reduces the competitive analysis of list accessing algorithms to lists of length two [1, 3, 5, 12, 20]. For example, this technique, in conjunction with phase partitioning, was used to prove that an algorithm called *TimeStamp* is 2-competitive [1, 3]. In this section, we repeat the development of this technique, but present it in a somewhat different way, in terms of a partial order on elements in the list. [7] This view leads us to a simple generalization of previous results and will assist us in our study of *WFA*′.

Consider the metrical task system corresponding to a list of length two. In this case there are two lists, (a,b) (a in front of b) and (b,a) (b in front of a), and the distance between these two states is 1. Since for all t we have $\omega_t((a,b)) - 1 \leqslant \omega_t((b,a)) \leqslant \omega_t((a,b)) + 1$, we can characterize the work functions at any given time t as having one of three distinct properties:

- $\omega_t((a,b)) < \omega_t((b,a))$, which we denote $a \succ b$,
- $\omega_t((a,b)) = \omega_t((b,a))$, which we denote $a \sim b$, or
- $\omega_t((a,b)) > \omega_t((b,a))$, which we denote $a \prec b$.

It is easy to verify directly from Eq. (1) the transitions between these three properties as a result of references in the string σ.

The resulting *three-state DFA* shown in Fig. 1 can be used to completely characterize the work functions, the optimal offline list configuration, and the optimal cost to service a request sequence σ. The start state of the DFA is determined by the initial order of the elements in the list: it is $a \succ b$ if the initial list is (a,b) and $a \prec b$ if the initial list is (b,a). Each successive request in σ results in a change of state in accordance with the transitions of the DFA, reflecting the work function values after serving that request.

Notice that the number of times a is referenced when in the state $a \prec b$ plus the number of times that b is referenced when in the state $a \succ b$ is equal to the total number of transitions into the middle DFA-state. The optimal sequence cannot avoid incurring cost upon such references. Therefore, the optimal cost of satisfying a sequence of requests σ is given by the number of transitions into the middle state of the DFA, plus the length of the sequence. The corresponding optimal offline strategy is: immediately before two or more references in a row to the same element, move that element to the front of the list.

Now consider list update for a list of length k. The cost of an optimal sequence can be written as the sum of the number of exchanges performed [8] and the reference costs

[7] This partial order has apparently been considered by Albers, von Stengel and Werchner in the context of randomized list update, and was used as a basis for an optimal randomized online algorithm for lists of length 4 [2].

[8] Recall that in our model we charge for each exchange, whether "paid" or "free"; each free exchange in the standard model precisely corresponds in our model to a reduced reference cost on the immediately following reference. See [15, Theorem 1].

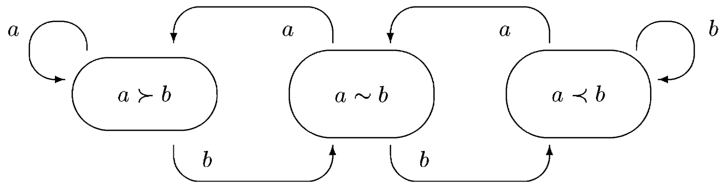

Fig. 1. The three-state DFA: the state $a \succ b$ corresponds to the case $\omega_t((a,b)) = \omega_t((b,a)) - 1$, the state $a \sim b$ corresponds to the case $\omega_t((a,b)) = \omega_t((b,a))$, and the state $a \prec b$ corresponds to the case $\omega_t((a,b)) = \omega_t((b,a)) + 1$.

at each state. For any pair of elements (a,b) we can identify a pairwise reference cost attributable to (a,b), adding one whenever b is referenced but a is in front of b in the list, or vice versa. The standard list factoring approach is to describe the cost of any optimal sequence for satisfying σ by decomposing it into $|\sigma|$ plus the sum over all pairs (a,b) of (i) this pairwise reference cost and (ii) all pairwise transpositions of a with b. For any pair (a,b), the sum of the pairwise transpositions and the pairwise reference cost describes a (possibly suboptimal) solution to the list of length two problem for the subsequence of σ consisting of references only to a and b. Therefore a lower bound on the optimal cost of satisfying σ is the sum of the costs of the optimal length-two solutions over all pairs (a,b), plus the length $|\sigma|$.

It is important to note that this "list factoring" lower bound is not tight.

Example 1. Consider a list of length five, initialized *abcde*, and the reference sequence $\sigma = ebddcceacde$. The sum of the length-two solutions, plus the length of σ, is 31; the optimal cost of satisfying σ is 32.

On the other hand, we do not know of any small examples where the optimal cost exceeds the list factoring lower bound by more than one, and we conjecture that the optimal cost does not exceed the lower bound by more than an additive constant related to the length of the list.

3.1. The partial order

We are thus led to consider the collection of $k(k-1)/2$ pairwise three-state DFAs, one for each pair a, b of elements in the list of length k. Consider the result of executing all these DFAs in parallel in response to requests in σ, starting from the states corresponding to the initial list. Fig. 2 shows an example. Each DFA defines a pairwise relation, $a \prec b$, $a \succ b$, or $a \sim b$ as the case may be, on the elements a and b. It is easy to verify that at every time t the resulting collection of relations defines a valid *partial order* on the k elements of the list. In particular, the list configuration obtained by following *Move-To-Front* at every step is always consistent with this partial order.

This partial order at each time t is defined by the reference sequence σ, and does not depend on any choice of algorithm for list update. When we refer to the "partial

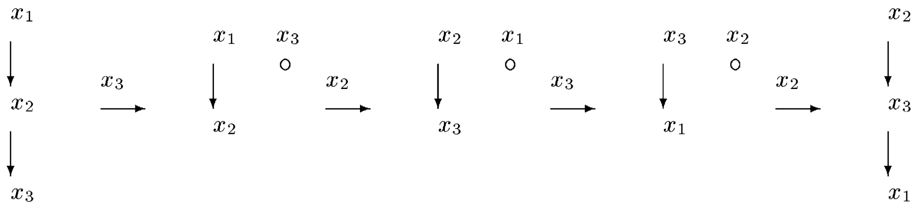

Fig. 2. Illustration of the evolution of the partial order on three elements in response to the request sequence $\sigma = x_3, x_2, x_3, x_2$ assuming the initial list is ordered x_1, x_2, x_3 from front to back. As usual, a directed edge from a to b indicates that $a \succ b$ in the partial order, whereas the absence of an edge indicates that $a \sim b$.

order", we mean this partial order as induced by a particular σ at a given time t. When we say that an algorithm is "consistent with the partial order", we mean that, when applied to a reference sequence σ, the list configuration visited by the algorithm at each time t, considered as a total order of the list elements, is consistent with the partial order induced by σ at that time t.

Define by G_t (respectively I_t) the number of elements greater than (respectively incomparable to) σ_t in this partial order immediately prior to its reference at time t. By the discussion above, the optimal cost of servicing a request sequence σ of length n and ending up in any state s is bounded below by the number of transitions into middle states of the DFAs, which at each step t is G_t. Hence for states s, $\omega_n(s) \geqslant n + \sum_{1 \leqslant t \leqslant n} G_t$.

An easy counting argument also shows:

Lemma 1.

$$\sum_t I_t \leqslant \sum_t G_t.$$

Proof. Since we start with a total ordering on the elements, determined by the initial ordering of the list, each two element DFA begins either in state $a \prec b$ or $a \succ b$. For each DFA, each transition out of its middle state $a \sim b$ must therefore be preceded by a transition into the middle state. Taken together, this implies that, cumulatively, the number of transitions out of middle states cannot exceed the number of transitions into middle states. Since $\sum G_t$ is the cumulative number of transitions into middle states of the DFAs, and $\sum I_t$ the cumulative number of transitions out of middle states, the result follows. $\square$

Lemma 1 leads to a useful characterization of online algorithms:

Theorem 1. *Any online list update algorithm that performs only free exchanges and maintains the invariant that the list order is consistent with the partial order is* $(2 - 1/k)$*-competitive.*

Proof. Any online algorithm A that maintains a list order consistent with the partial order and performs no paid exchanges has a total cost $A(\sigma)$ satisfying $A(\sigma) \leqslant n + \sum_t (I_t + G_t)$, where $|\sigma| = n$.

By Lemma 1 and the fact that $OPT(\sigma) \leqslant kn$, we can conclude that $A(\sigma) \leqslant n + 2\sum_t G_t \leqslant (2 - 1/k)OPT(\sigma)$. □

3.2. Competitive analysis of online algorithms

Theorem 1 provides a new, simple proof that a collection of online algorithms (many already known to be competitive) are all $(2-1/k)$-competitive. These algorithms include *Move-To-Front*, *TimeStamp*, *MRI*(i), and *SBR*(α) [1, 11, 17]. Each of these online algorithms moves only the referenced element. By Theorem 1, it is enough to show that these algorithms maintain lists consistent with the partial order.

We observed above that *Move-To-Front* maintains lists consistent with the partial order. Suppose the list is consistent with the partial order at time t, immediately before a reference to x. Then immediately after the reference (and after x is moved to the front), each element of the list is less than or incomparable to x, and is also behind x in the list. And because the respective pairwise order of other elements does not change, the list remains consistent with the partial order at time $t + 1$.

The *TimeStamp* algorithm (originally called *TimeStamp*(0)) due to Albers [4] is defined as follows:

> On a request for an item x, insert x in front of the first (from the front of the list) item y that precedes x on the list and was requested at most once since the last request for x. Do nothing if there is no such item y or if x is being requested for the first time.

The *TimeStamp* algorithm makes only free exchanges. Furthermore, by construction, after a reference to x, each item y that precedes it in the resulting list must have been requested at least twice since the last request for x. Therefore every element in front of x is incomparable to x (and not less than x) after the request. Each element behind x is less than or incomparable to x. Finally, the respective orders of other elements do not change as a result of the reference to x. Immediately prior to the initial reference to x, all elements in front of it are greater than it in the partial order. Hence *TimeStamp* —and indeed any algorithm that moves x forward at least as far as *TimeStamp* does— maintains a list order consistent with the partial order.

Ran El-Yaniv has recently presented another family of algorithms, the *MRI*(k) family [11]:

> On a request for an item x, move x forward to just behind the rearmost item y that precedes x on the list and was requested at least $k + 1$ times since the last request for x. If there is no such item y or if x is being requested for the first time, move x to the front.

El-Yaniv shows that *MRI*(1) is equivalent (except for the first move of each element) to *TimeStamp*. Because any element that is requested more than twice since the last reference to x must be incomparable to x after the reference to x, Theorem 1 applies to *MRI*(i) for all i.

Schulz has recently presented the *SBR*(α) family [17]. From his Lemma 1 and the definition, the referenced element is moved forward at least as far as *TimeStamp*. As shown above, any such algorithm maintains a list order consistent with the partial order.

We have shown:

Corollary 1. *Move-To-Front*, *TimeStamp*, *MRI*(i) *and* *SBR*(α) *are all* $(2 - 1/k)$-*competitive.*

4. On the performance of work function algorithms

4.1. Preliminaries

We begin with some definitions and facts. In what follows, the $(t + 1)$st request σ_{t+1} is x. The task cost $\tau(s)$ is denoted $x(s)$, which is the depth of x in the list configuration s. As before, we denote by s_t the state visited by the work function algorithm at time t, immediately before servicing the request to x.[9]

We first define the $\uparrow_x$ binary relation on two states.

Definition 6. $s \uparrow_x s'$ iff s and s' are identical, or if s' can be derived from s by moving x forward while leaving the relative positions of other elements undisturbed.

Where x is understood from context, we write simply $s \uparrow s'$. In the case of list update, the "free exchange" cost model (or equivalently, the distance based on interchanges) implies that whenever $s \uparrow_x s'$, $x(s) = x(s') + d(s, s')$. This property in turn implies the following property for wfa-eligible states:

Proposition 7. *Suppose s is wfa-eligible, and $s \uparrow_x s'$. Then $\omega_{t+1}(s') \leqslant \omega_{t+1}(s)$. (Moving x forward cannot increase the work function.) Furthermore, s' is wfa-eligible, and indeed $\omega_{t+1}(s) = \omega_{t+1}(s')$.*

Proof. We start with the first half of the statement, considering first s fundamental, and then more generally s wfa-eligible. Suppose first that s is fundamental, $\omega_{t+1}(s) = \omega_t(s) + x(s)$. We have $\omega_{t+1}(s') \leqslant \omega_t(s') + x(s')$ by Proposition 3, and $\omega_t(s') \leqslant \omega_t(s) + d(s', s)$ by Proposition 1, so $\omega_{t+1}(s') \leqslant \omega_t(s) + d(s', s) + x(s')$. But $d(s', s) + x(s') = x(s)$ so $\omega_{t+1}(s') \leqslant \omega_t(s) + x(s) = \omega_{t+1}(s)$ as was to be shown.

Now suppose more generally that s is wfa-eligible. By Proposition 1, we have $\omega_{t+1}(s) = \omega_{t+1}(f) + d(f, s)$ for some fundamental state f, for which also $x(f) = x(s)$.

[9] Again, we do not distinguish between a reference to x followed by free exchanges, on the one hand, and "paid" exchanges moving x forward, followed by a lower-cost reference to x, on the other. We refer to either of these combined steps as "servicing" the request to x.

This means that $\omega_{t+1}(s)=\omega_t(f)+d(f,s)+x(f)=\omega_t(f)+d(f,s)+x(s)$. But $\omega_{t+1}(s') \leqslant \omega_t(s')+x(s') \leqslant \omega_t(f)+d(f,s')+x(s') \leqslant \omega_t(f)+d(f,s)+d(s,s')+x(s')=\omega_t(f)+d(f,s)+x(s)$, so $\omega_{t+1}(s') \leqslant \omega_{t+1}(s)$.

Next, we show the stronger properties in the second half of the statement. Since $d(s',s_t) \leqslant d(s',s)+d(s,s_t)$, we also have $\omega_{t+1}(s')+\tau(s')+d(s',s_t) \leqslant \omega_{t+1}(s)+\tau(s)+d(s,s_t)$. This means that s' is also wfa-eligible. If the inequality $\omega_{t+1}(s') \leqslant \omega_{t+1}(s)$, were strict, s could not be wfa-eligible. So we have specifically $\omega_{t+1}(s')=\omega_{t+1}(s)$. □

Recall from Proposition 6 that $\tau(s) \leqslant \tau(s_t)$, so the work function algorithm cannot move x backward.

We can now show that (a) there always exists a wfa-eligible state that requires no paid exchanges, and (b) that if *WFA′* is restricted to moving the referenced element only, it is equivalent to the following algorithm ("*Move-To-Min-ω*"):

Mtmω: On a reference to x, move x forward (or not at all) to a state with lowest work function value immediately after the reference.

In other words, if s_t is the state the algorithm is in immediately before servicing the $(t+1)$st request σ_{t+1}, then *Mtmω* moves to a state s_{t+1} for which $s_{t+1}=\operatorname{argmin}_{(s\,:\,s_t \uparrow_x s)}\omega_{t+1}(s)$, and satisfies σ_{t+1} there. Summarizing:

Proposition 8. *Mtmω is a special case of WFA′ and* Move-To-Front *is a special case of Mtmω.*

Proof. We first show that *Mtmω* is a special case of *WFA′*. That is, we need to show that any state produced by *Mtmω* is wfa-eligible. Suppose s is such a state for which $s_t \uparrow s$, and s minimizes $\omega_{t+1}(s)$ among all such. Let s' be *some* wfa-eligible state for which $s_t \uparrow s'$. (The existence of such a state is demonstrated below.) Then either $s' \uparrow s$ or $s \uparrow s'$. If the former, Proposition 7 applies and s is wfa-eligible. If the latter, $d(s,s')=(x(s)-x(s'))$ and $\omega_{t+1}(s) \leqslant \omega_{t+1}(s')$ together imply that $\omega_{t+1}(s)+x(s)+d(s_t,s) \leqslant \omega_{t+1}(s')+x(s')+d(s_t,s')$, and again s is wfa-eligible.

It remains to demonstrate that there is at least one wfa-eligible state s for which $s_t \uparrow s$. For convenience in what follows, we denote generally by $\hat{s}$ the state formed from s by moving x to the front without changing the order of other elements, $s \uparrow_x \hat{s}$ and $x(\hat{s})=1$. We show that the move-to-front state $\hat{s}_t$, the state which simultaneously satisfies $s_t \uparrow \hat{s}_t$ and $x(\hat{s}_t)=1$, is wfa-eligible. By Proposition 7, there must be some r wfa-eligible for which $x(r)=1$ (for any wfa-eligible r', take $\hat{r}'$). It is a basic fact of permutation distance that $d(r,s_t)=d(r,\hat{s}_t)+d(\hat{s}_t,s_t)$, because the interchanges in $d(r,s_t)$ not involving x can all be resolved first, without moving x. Given this fact, then $\omega_{t+1}(r)+x(r)+d(r,s_t)=\omega_{t+1}(r)+x(\hat{s}_t)+d(r,\hat{s}_t)+d(\hat{s}_t,s_t)$. But $\omega_{t+1}(r)+d(r,\hat{s}_t) \geqslant \omega_{t+1}(\hat{s}_t)$ by Proposition 1, hence $\omega_{t+1}(\hat{s}_t)+x(\hat{s}_t)+d(s_t,\hat{s}_t) \leqslant \omega_{t+1}(r)+x(r)+d(s_t,r)$, which was to be proved.

As a corollary, the algorithm *Move-To-Front* is a special case of the work function algorithm. □

4.2. WFA' is $O(1)$-competitive for list update

In the preceding section, we characterized the work function algorithm in terms of the work function values of states formed by moving the referenced element forward. We noted that the work function value cannot increase as the referenced element is moved forward. In order to prove results about the work function algorithm, however, we must characterize all states to which the work function algorithm could move; and thus we must characterize circumstances under which the work function value must *strictly decrease*. Our proof technique, then, supposes by hypothesis that the work function algorithm encounters a state of a particular undesired type; we consider the optimal sequence of interchanges and references that leads to the given work function value; then we must construct a new sequence, leading to a state identical to the first but for moving the referenced element forward, for which the total cost (of references and interchanges) is strictly lower.

The technically challenging part of the proof is the following lemma.

Lemma 2. *Consider $\sigma=\sigma_1,\ x,\ \sigma_2,\ x$, where in σ_2 there are no references to x, and $|\sigma|=t$. Let S be any fundamental state at the final time step t.*

Let $\mathcal{N}$ be the set of elements that are not referenced in σ_2 that are in front of x in S, and let $\mathcal{R}$ be the set of elements (not including x) that are referenced in σ_2. Also, let $\hat{S}$ be S with x moved forward just in front of the element in $\mathcal{N}$ closest to the front of the list. Then

$$\omega_t(\hat{S}) \leqslant \omega_t(S)+|\mathcal{R}|-|\mathcal{N}|. \tag{3}$$

Proof. Suppose O is an optimal sequence of pairwise interchanges punctuated by references, ending in the state S after satisfying the entire reference sequence $\sigma=\sigma_1,x,\sigma_2,x$. Then the cost of O is the work function value $\omega_t(S)$. Let T denote the state in which O satisfies the penultimate reference to x (that between σ_1 and σ_2). We note that, at the point immediately prior to the penultimate reference to x (at time u, say), the cost of O up to that point is $\omega_{u-1}(T)$. In this construction, we modify O between T and S so as to obtain the state $\hat{S}$, with $S\uparrow_x\hat{S}$ and $\omega_t(\hat{S})\leqslant\omega_t(S)-|\mathcal{N}|+|\mathcal{R}|$.

Let N denote the total number of elements *not* referenced between $\sigma_u=x$ and $\sigma_t=x$. (This set specifically includes x, and is potentially much larger than $|\mathcal{N}|$, which is the number of such elements in front of x in S.) Label these nonreferenced elements $p_1,\ldots,p_N$ in the order they occur in the state T, with p_1 referring to the one such closest to the front of the list.

Denote by $I[X,Y]$ the number of interchanges of nonreferenced elements (other than x) in a given sequence between the states X and Y.

The construction of the lower-cost state $\hat{S}$ proceeds in three stages (see below for a diagram):

1. Rearrange the respective order of the nonreferenced elements within T to obtain some state T'. In T', x will occupy the location of the front-most nonreferenced element in T. All other nonreferenced elements p in T' will satisfy a *nondecreasing depth*

property, that $p(T) \leqslant p(T')$.[10] All referenced elements remain at their original depths. (The specific definition of the state T' will emerge from the rest of the construction; the cost of the modified sequence can be bounded by using only the nondecreasing depth property.) Evaluate $\sigma_u = x$ in this state T'.

Using the nondecreasing depth property, we show (Proposition 9, proof deferred to the appendix) that $x(T') + d(T, T') \leqslant x(T) + |\mathscr{R}| + I[T, T']$ (where $I[X, Y]$ is defined as above).

2. Considering the portion of O beginning immediately after the penultimate reference to x at time u, as a sequence of pairwise interchanges transforming T to S, $O : T \to S$, apply a suitably chosen subsequence O', including all of the references and many of the transpositions, of O. This subsequence O' will transform T' to a state S'. In this state S', (i) each referenced element has the same depth as it does in S; (ii) the element x occupies the position of the front-most nonreferenced element in S; and (iii) all other nonreferenced elements in S' are in their same respective pairwise order as in S. Evaluate x in S'.

We show (Proposition 11, proof deferred to the appendix) that such a transformation from some T' with the nondecreasing depth property, to S' as so defined, can be achieved by a suitably chosen subsequence of O. We also show that $I[T, T'] + I[T', S'] \leqslant I[T, S]$, by showing that the interchanges between nonreferenced items in the transformations from T to T' and from T' to S' are all contained in O.

3. Transform S' to the state $\hat{S}$, where $\hat{S}$ is defined by (i) $S \uparrow_x \hat{S}$, and (ii) the depth of x in $\hat{S}$ is the depth of the front-most nonreferenced element in S (which is also its depth in S').

We show (Proposition 10, proof deferred to the appendix) that $x(S') + d(S', \hat{S}) + |\mathscr{N}| \leqslant x(S)$.

This process can be illustrated as follows, using $\rightarrow$ to denote a reference, and $\rightsquigarrow$ to denote pairwise interchanges between references. The original, hypothetically optimal sequence O can be depicted as:

$$\cdots \rightsquigarrow T \xrightarrow{x} T \rightsquigarrow \cdots \rightsquigarrow S \xrightarrow{x} S.$$

(Recall that we assume that O satisfies $x = \sigma_t$ in S.)

After the above modifications (denoted 1, 2, 3), the modified sequence O' is

$$\ldots \overset{O}{\rightsquigarrow} T \overset{1}{\rightsquigarrow} T' \xrightarrow{x} T' \ldots \overset{2}{\rightsquigarrow} S' \xrightarrow{x} S' \overset{3}{\rightsquigarrow} \hat{S}.$$

The result now follows by comparing the cost of the modified sequence to the cost of the original sequence, from and after T (both sequences incur $\omega_{u-1}(T)$ to that point). The cost attributable to the original sequence is the sum of

1. $x(T)$;
2. the cost of references in σ_2;
3. from T to S, the cost of interchanges between referenced elements;

[10] Recall that we denote the depth of an element p in the state X by $p(X)$.

4. from T to S, the cost of interchanges between a referenced and a nonreferenced element;
5. from T to S, the cost $I[T,S]$ of interchanges between nonreferenced elements; and
6. $x(S)$.

The cost attributable to the modified sequence O' is the sum of
1. from T to T', the cost of all interchanges;
2. $x(T')$;
3. the cost of references in σ_2;
4. from T' to S', the cost of interchanges between referenced elements;
5. from T' to S', the cost of interchanges between a referenced and a nonreferenced element;
6. from T' to S', the cost $I[T',S']$ of interchanges between nonreferenced elements;
7. $x(S')$; and
8. from S' to $\hat{S}$, the cost of all interchanges.

By construction, items two, three and four for the sequence O are identical to items three, four and five for the sequence O'. Thus we compare $x(T) + I[T,S] + x(S)$ for the first sequence to $d(T,T') + x(T') + I[T',S'] + x(S') + d(S',\hat{S})$ for the second sequence.

Given $x(T') + d(T,T') \leqslant x(T) + |\mathscr{R}| + I[T,T']$ (Proposition 9), $I[T,T'] + I[T',S'] \leqslant I[T,S]$ (Proposition 11), and $x(S') + d(S',\hat{S}) + |\mathscr{N}| \leqslant x(S)$ (Proposition 10), the result follows by substitution. □

We obtain the following corollary to Lemma 2.

Corollary 2. *Consider a request sequence σ where the last request (the tth request in σ) is to x. If s is wfa-eligible after executing σ, then the depth of x in s is at most $2|\mathscr{R}|$, where $\mathscr{R}$ is the set of elements that have been referenced since the penultimate reference to x.*

Proof. Let f be a fundamental state such that $\omega_{t+1}(s) = \omega_{t+1}(f) + d(f,s)$. By Proposition 5, f is also wfa-eligible and $x(f) = x(s)$. Suppose $x(s) > 2|\mathscr{R}|$. Then $x(f) > 2|\mathscr{R}|$. Elements in front of x in f either have or have not been referenced since the penultimate reference to x; so $x(f) > 2|\mathscr{R}|$ implies $|\mathscr{N}| > |\mathscr{R}|$, where $\mathscr{N}$ is the set of elements in front of x in f that have not been referenced since the penultimate reference to x. Then by Lemma 2 there exists $\hat{f}$ with $\omega_t(\hat{f}) < \omega_t(f)$ and $f \uparrow_x \hat{f}$, contradicting the assumption that f is wfa-eligible. □

Finally, we use the lemma to obtain the main theorem.

Theorem 2. *WFA′ is $O(1)$-competitive.*

Proof. Consider an arbitrary element x, and let $\sigma = \sigma_0, x, \sigma_1, x, \sigma_2, x$, where in σ_1 and σ_2 there are no references to x. Then by Lemma 2 and Corollary 2 the depth of

x in the $Mtm\omega$ state, immediately before the final reference to x, is at most $2r_1 + r_2$, where r_1 is the number of distinct elements referenced in σ_1 and r_2 is the number of distinct elements referenced in σ_2, not referenced in σ_1, that are moved in front of x at some point during the subsequence σ_2.

As usual, let G be the number of elements greater than x immediately before its final reference and let I be the number of elements incomparable to x immediately before its final reference. In addition, let L be the number of elements less than x immediately before its final reference that were incomparable to x immediately before the penultimate reference to x. We have $r_1 + r_2 \leqslant G + I + L$.

Denote by t_1 the time of the penultimate reference to x, and by t_2 the time of the final reference. Since each element in L at time t_2 is incomparable to x at time t_1, we have $L_{t_2} \leqslant I_{t_1}$. That is, for any t_2, there is some $t_1 < t_2$ such that $L_{t_2} \leqslant I_{t_1}$. Thus $\sum_t L_t \leqslant \sum_t I_t$. But $\sum_t I_t \leqslant \sum_t G_t$ by the counting argument, Lemma 1. Summarizing, we have

$$\begin{aligned} WFA'(\sigma) &\leqslant \sum_t (2r_1 + r_2) \leqslant 2(r_1 + r_2) \\ &\leqslant 2\sum_t (G_t + I_t + L(0)_t) \leqslant 6\sum_t G_t \leqslant 6OPT(\sigma). \end{aligned}$$

Therefore, WFA' (equivalent to $Mtm\omega$, as we have defined it) is at least 6-competitive. □

Note that, for list update, the algorithm *WFA* (without the prime) can in some circumstances be less effective than WFA'. Consider the sequence $\sigma = bbb$ for a two-element list (a, b). After the second reference to b, the list configuration (b, a) has strictly lower work function value. But *WFA* does not (necessarily) move to that state until after the *third* reference to b. Nevertheless, as noted above it is possible (by expanding the construction of the DFA to more states) to extend the above proof of $O(1)$-competitiveness to *WFA*.

It is fairly clear that the competitive ratios shown by our analyses of these algorithms are not tight. A generalization of the above example to longer lists shows that *WFA*, even without paid exchanges, is no better than 3-competitive.

Acknowledgements

We gratefully acknowledge discussions with Susanne Albers, Ran El-Yaniv, Sandy Irani, and T.S. Jayram. The detailed comments of our anonymous reviewers considerably clarified the presentation.

This work was supported in part by NSF grant EIA-9870740 and BSF grant 96-00247 (Karlin), the CRA Distributed Mentor Project (Hildrum and Rasala), and an IBM Research Fellowship (Anderson).

Appendix

In this appendix, we address the proofs of the three propositions leading to Lemma 2. The most intricate part of the part of the construction is contained in Proposition 11; we save its proof for last.

Proposition 9. *Let the elements of the list T be divided into two classes $\mathscr{R}$ and $\mathscr{N}$, referred to in the text as the "referenced" elements and the "nonreferenced" elements, respectively. (The meanings of "referenced" and "nonreferenced" are not used in the proofs of this and the next proposition.) Let x be a designated element from the "nonreferenced" class.*

Suppose the list ordering T' is derived from T as follows:

- *All referenced elements $r \in \mathscr{R}$ have $r(T') = r(T)$;* [11]
- *The element x occupies in T' the location of the front-most nonreferenced element in T; and*
- *T' has the nondecreasing depth property, $p(T') \geqslant p(T)$ for all $p \neq x$, $p \in \mathscr{N}$.*

Then the number of pairwise interchanges required to transform T to T' is no greater than

- *the number of interchanges involving x, plus,*
- *the number of referenced elements $|\mathscr{R}|$, plus,*
- *the number of interchanges involving only nonreferenced elements $p, q \in \mathscr{N}$, p, $q \neq x$.*

In particular, the depth of x in T is equal to the depth of x in T', plus the number of interchanges involving x.

Proof. If the front-most nonreferenced element in T is already x itself, then in order to satisfy the nondecreasing depth property, we must have $p(T') = p(T)$ for all nonreferenced elements, and there is nothing to prove.

Otherwise, denote by $p_1, \ldots, p_N$ the nonreferenced elements in their order in T, with the designated element $x = p_z$ for some z. We note that for elements behind x, p_i, $i > z$, the nondecreasing depth property requires $p_i(T) = p_i(T')$. Thus the rearrangement from T to T' is limited to x and elements in front of x.

We construct a sequence of interchanges from T to T' as follows. Begin by moving x forward to the location of the first nonreferenced element, p_1. These interchanges all involve x. Next, we move p_1 backwards to location 2. [12] By the nondecreasing depth property, either p_1 or p_2 must occupy location 2 in T'. If p_2 is x, then p_1 must occupy location 2, and we are done. Otherwise, consistent with the nondecreasing depth property, there may be an interchange between p_1 and p_2. Whichever of p_1 and p_2 is *not* in its correct location in T' we move backwards to location 3.

[11] Recall that we denote the depth of r in T by $r(T)$.

[12] In what follows, by slight abuse of notation, we refer to the location of the i'th nonreferenced element in T by the description "location i" or "position i".

Inductively, at step i, for (nonreferenced) location i, we have

- each referenced element in front of location i has interchanged with at most one nonreferenced element;
- each referenced element behind location i has not interchanged with any nonreferenced elements; and
- some element $p_j, j<i$, is in location i, and either (i) p_i is x; or (ii) p_j is immediately adjacent to p_i.

If p_i is x, we are done. Otherwise, by the nondecreasing depth property one or the other of p_i and p_j must occupy location i. We swap p_i and p_j if necessary; and continue by moving the rearward of the two toward location $i+1$. By induction, when this process completes, each referenced element has interchanged with at most one nonreferenced element (other than x), and the result follows. □

We next prove Proposition 10, which presents a construction that is in some sense the obverse of that in Proposition 9. At this point in the main proof, x is already ahead of all other nonreferenced elements. We must move the nonreferenced elements *forward* to their ending positions, so that they occupy the same positions (but for x) as in the hypothetically optimal state.

Proposition 10. *Let x be a designated element of the list S, and let the remaining elements of the list S be divided into two classes $\mathscr{R}$ and $\mathscr{N}$. These classes are referred to in the text as the "referenced" elements $\mathscr{R}$ and, together with x, the "nonreferenced" elements $\mathscr{N}\cup\{x\}$, respectively. (The meanings of "referenced" and "nonreferenced" are not otherwise needed in the proof of this proposition.) In what follows, we ignore all elements p of depth greater than x in S, $p(S)>x(S)$; henceforth we can assume that the "nonreferenced" class $\mathscr{N}$ includes only elements in front of x in S. (This assumption conforms to the usage in the text.)*

Suppose the list ordering S' is derived from S as follows:

- *All referenced elements $r\in\mathscr{R}$ occupy the same locations in the same order in S' as in S.*
- *The designated element x occupies in S' the position of the front-most nonreferenced element in S. (For all $p\in\mathscr{N}\cup\{x\}$, $x(S')\leqslant p(S)$.)*
- *All other nonreferenced elements are in the same pairwise order in S' as in S. (For all $p,q\in\mathscr{N}$, $p,q\neq x$, $p(S)<q(S)\Leftrightarrow p(S')<q(S')$.)*

Suppose the list ordering $\hat{S}$ is derived from S by moving x forward to immediately in front of the front-most nonreferenced element, but making no other interchanges.

Then the cost $x(S')$ of the reference to x in state S', plus the distance $d(S',\hat{S})$ to transform S' to $\hat{S}$, is less than the depth $x(S)$ of x in S by at least the number of nonreferenced elements in front of x in S. That is, we have

$$x(S')+d(S',\hat{S})+|\mathscr{N}|\leqslant x(S).$$

Proof. Suppose x occupies the i'th nonreferenced location from the front in S. (That is, suppose $|\mathscr{N}|=i-1$.) Denote the first i nonreferenced elements of S in order by

$q_1,\dots,q_i = x$. In S', the element q_{i-1} occupies position i; q_{i-2}, position $i-1$; and so on; q_1 occupies position 2; and x occupies position 1. We transform S' to $\hat{S}$ by interchanging, for all $1<j<i$, q_j with all referenced elements between it and q_{j-1}, and q_1 with all referenced elements between it and x. Each referenced element between x and q_{i-1} interchanges with at most one nonreferenced element, and each such is in front of x in S. Thus the number of exchanges required to transform S' to $\hat{S}$, plus the number of referenced elements in front of x in S', plus the number of nonreferenced elements in front of x in S, is no greater than the depth of x in S. The result follows. □

Finally, we address the most intricate part of the construction:

Proposition 11. *Let the reference sequence $\sigma = \sigma_1, x, \sigma_2, x$, where in σ_2 there are no references to x. Consider an arbitrary sequence of pairwise interchanges and references that satisfies σ. Denote by T the list ordering obtained by that sequence immediately prior to the penultimate reference to x, and by S the list ordering immediately after the final reference to x. Let O denote the sequence of interchanges that transforms T to S, writing $O: T \rightsquigarrow S$ or $O(T) = S$, and $|O|$ the number of interchanges in the sequence. Let $\mathscr{R}$ (respectively, $\mathscr{N}$) denote the list elements that are referenced (respectively, not referenced) by σ_2, that is, referenced (or not) between T and S. For convenience, designate x as a "nonreferenced" element unless otherwise indicated.*

Suppose S' is derived from S, with the properties that
- *all referenced elements $r \in \mathscr{R}$ are in the same position, $r(S') = r(S)$;*
- *x occupies in S' the position of the front-most nonreferenced element in S, $x(S') \geqslant p(S)\ \forall p \in \mathscr{N}$; and*
- *all other nonreferenced elements $p, q \in \mathscr{N}, p, q \neq x$ are in their same respective order in S' as in S, $p(S) < q(S) \Leftrightarrow p(S') < q(S')$.*

Then there is a T' with the nondecreasing depth property, $p \in \mathscr{N}, p \neq x \Rightarrow p(T') \geqslant p(T)$, and a subsequence $O' \subseteq O$ of interchanges, such that
- *O' transforms T' to S', $O'(T') = S'$; and*
- *the number of interchanges in O is at least the number in O' plus the number $I[T, T']$ of interchanges of nonreferenced elements (other than x) necessary to derive T' from T, $|O| \geqslant |O'| + I[T, T']$.*

Proof. As in the proof to Proposition 9, we denote by p_i the nonreferenced element occupying the i'th nonreferenced position in T, with $x = p_z$ for some z.[13] Throughout the construction, the location of referenced elements $r \in \mathscr{R}$ remains fixed, and we focus instead on the $N = |\mathscr{N}|$ positions of nonreferenced elements.

[13] We use the terms "position" and "location" interchangeably to refer specifically to the respective positions of *nonreferenced* elements in T.

First, consider O as a sequence of interchanges of numbered elements, $O: T \rightsquigarrow S$, and consider its inverse, $O^{-1}: S \rightsquigarrow T$. We start with $T'_{\text{init}} = O^{-1}(S')$, the list ordering obtained by applying the interchanges in O, in reverse order, to S'. By progressively removing interchanges from O, we eventually obtain a $T' = (O')^{-1}(S')$ for which the nondecreasing depth property holds. Along the way, we demonstrate that all transpositions of nonreferenced elements required to get from T to T', $I[T, T']$, are accounted for as interchanges removed from O. The result then follows.

We construct T' iteratively beginning with $T'_{\text{init}} = O^{-1}(S')$, and beginning at the rear of the list. For convenience, we describe the iteration as proceeding from $i = N$, the final nonreferenced element, to $i = 1$, the front-most nonreferenced element. (The base case is denoted by "$i = N + 1$".) At each step, then, we define a map $O_i: T'_i \to S'$, which is a subsequence of O. The nondecreasing depth property is maintained for those elements (other than x) in $O_i^{-1}(S') = T'_i$ that occupy the locations i through N in T'_i. We further show that any necessary interchanges of elements as we proceed from T'_i to T'_{i-1} correspond to transpositions in O'_i.

For each pair of elements $p, q \neq x$ at locations i and greater in T'_i, we can determine whether these two elements are in the same or in the opposite order in T. We denote by $I_i[T, T'_i]$ the number of pairwise inversions of such elements (other than x). We denote by $|O|$ (respectively, $|O_i|$) the number of transpositions in the sequence O (respectively, O_i).

Formally, we show by induction that for each i:

1. $O_i(T'_i) = S'$ (and $O_i^{-1}: S' \to T'_i$)
2. $O_i \subseteq O$ in the sense of a subsequence of transpositions, and $|O| \geqslant |O_i| + I_i[T, T'_i]$ (all swaps and inversions are accounted for)
3. $x(T'_i) \leqslant p_i(T)$ (x is no deeper than position i)
4. $\forall p \neq x$ with $p(T) \geqslant p_i(T)$, $p(T'_i) \geqslant p(T)$ (all elements at position i or greater in T, other than x, have the nondecreasing depth property)
5. $\forall p, q \neq x$ with $p(T), q(T) < p_i(T)$:
   (a) $p(S) < x(S) \Leftrightarrow p(T'_i) \neq p(T)$, and $p(S) > x(S) \Leftrightarrow p(T'_i) = p(T)$
   (b) $p(T) = q(T'_i) \Rightarrow p(S) > q(S)$.

To carry out the induction proof, we will start by demonstrating the hypotheses for an appropriate base case. For the induction step, we assume the five hypotheses for $i+1$, derive a transformation O_i, and show the validity of the hypotheses for i. Then we define $T' = T_1$, and note that the nonincreasing depth property is satisfied for all $p \neq x$. We define $O' = O_1$, and note all of the inversions between nonreferenced elements in T' have been accounted for, i.e., $I[T, T'] + |O'| \leqslant |O|$. Finally, we repeat that because the only transpositions removed from O are between nonreferenced elements, the depths, and thus the reference costs, of all referenced elements remains identical between O and O'. Thus O' can be extended to a sequence of transpositions and references whose cost on σ differs from that of O only by the number of transpositions incurred and by the cost of the penultimate and final references to x. The result in the text then follows.

The base case. For the base case, we define $O_b = O$, $T'_b = O^{-1}(T')$. (Notationally, b is $N + 1$.) Then (1) and (2) follow from our definition (I_b is zero). Items (3) and (4) are vacuous. We must show that items (5)(a) and (5)(b) are true for all nonreferenced elements in T'_b. For item (5)(a), by construction of S', elements q deeper than x in S are unaffected by the shift, $q(S)>x(S) \Rightarrow q(S)=q(S')$ so $q(T)=q(T'_b)$, while elements q closer to the front of S are "shifted down", $q(S)<x(S) \Rightarrow q(S') \neq q(S)$ (indeed $q(S')>q(S)$), so $q(T) \neq q(T'_b)$.

Finally, for (5)(b), $p(T)=q(T'_b), p \neq q$ implies $p(T) \neq p(T'_b)$ implies (by (5)(a)) $p(S)<x(S)$, similarly $q(S)<x(S)$. By construction, $p(S')>p(S)$ by one nonreferenced position, but $q(S')=p(S)$ since O^{-1} takes q to location $p(T)$ in T'_b. Hence $p(S')>q(S')$.

Induction step. Now suppose the entire hypothesis is true for $i+1$ (including for example $b=N+1$). We will construct an appropriate mapping O_i that satisfies the hypotheses for i. We describe three stages, depending on whether the element x has yet been considered. Denoting by z the location of x in T, so that $x=p_z$, we consider p_i for (i) $i>z$, (ii) $i=z$, (iii) $i<z$ in turn.

Case (i): $i>z$. By induction, we have the nondecreasing depth property for all elements $p_j, j \in i+1,\ldots,N$. Because $i>z$, this requires strictly that $p_j(T'_{i+1})=p_j(T)$ for all such locations j. In this case, the nondecreasing depth property applied to i will require strictly that $p_i(T)=p_i(T')$. Throughout this stage, in particular, the nondecreasing depth property for i implies $I_i[T,T'_i]=0$.

We examine the current occupant of position i in T'_{i+1}. There are three possibilities to consider:

- The occupant is p_i itself, $p_i(T'_{i+1})=p_i(T)$. In that case, set $O_i=O_{i+1}$. The nondecreasing depth property is (precisely) satisfied. Hypotheses (1) and (2) follow immediately from their validity for $i+1$; hypothesis (3) and (4) follow from the depth property; and hypothesis (5) is more restrictive, hence valid.
- The occupant is x, $x(T'_{i+1})=p_i(T)$. In this case, T'_i will be obtained by interchanging p_i with x. We observe that $p_i(T'_{i+1})<x(T'_{i+1})$ (by the depth property at $i+1$), and $x(S')<p_i(S')$ by construction (x is the front-most nonreferenced element in S'), so x and p_i are inverted by O_{i+1}, and there is a transposition in O_{i+1} between them. Remove this transposition to get $O_i \subset O_{i+1}$. The nondecreasing depth property is again precisely satisfied, implying hypotheses (3) and (4); and hypothesis (5) is unchanged for $p_j, j<i$. (Hypothesis (5) does not apply to x.)
- The occupant is $p_j \neq x$, $p_j(T'_{i+1})=p_i(T)$. In this case, T'_i will be obtained by interchanging p_j with p_i. We observe (again) that the depth property is precisely satisfied for $k>i$, so $p_i(T'_{i+1})<p_j(T'_{i+1})=p_i(T)$. Also, by hypothesis (5)(b), we have $p_j(T'_{i+1})=p_i(T) \Rightarrow p_j(S)<p_i(S)$. Therefore, O_{i+1} inverts p_i and p_j. Remove this transposition to get $O_i \subset O_{i+1}$. The nondecreasing depth property is again precisely satisfied, implying hypotheses (3) and (4). We note that hypothesis (5) holds for p_j by transitivity: suppose p_i occupies location k in T'_{i+1}, $p_k(T)=p_i(T'_{i+1})$; and $p_i(T)=p_j(T'_{i+1})$; so by hypothesis (5)(b) at the previous step we have $p_j(S)<p_i(S)<p_k(S)$, implying (5)(b) at the current step (and in particular $j \neq k$, so (5)(a) is satisfied).

This concludes the analysis of the stage where the depth property is precisely satisfied, $p_i(T') = p_i(T)\ \forall i > z$.

Case (ii): $i = z$. In this case there is nothing to do, $O_i = O_{i+1}$. Since the depth property is satisfied for $i+1,\dots,N$, and does not apply to x, it remains satisfied.

Case (iii): $i < z$. In this case, the elements occupying locations $i+1,\dots,N$ are $p_{i+1},\dots,p_N$ other than x, together with (by the nondecreasing depth property) some single element p_j, which might be x.

We consider four cases: p_j is x; p_j is p_i; $p_j \neq x$, and p_i occupies location i in T'_{i+1}; $p_j \neq x, p_i$, and $p_i(T'_{i+1}) \neq p_i(T)$.

- p_j is x. In this case, by hypothesis (3), $p_j = x$ occupies location $i+1$ in T'_{i+1}. In this case, T'_i will be obtained by interchanging p_i and x. We know p_i is above x in T'_{i+1}, $p_i(T'_{i+1}) < x(T'_{i+1})$. But, as above, $x(S') < p_i(S')$ by construction (x is the front-most such in S'). Therefore there is a swap in O_{i+1} between x and p_i. Remove it to obtain $O_i \subset O_{i+1}$. We note that $I_i = I_{i+1}$, because x occupied location $i+1$ (hypothesis (3)), and all elements below location $i+i$ have index larger than i. The occupant of location i is left undetermined here; even if it is x, hypothesis (3) remains satisfied. Hypothesis (5) remains true, since only x's location has changed, and the hypothesis does not apply to it.
- p_j is p_i. In this case, we do nothing, $O_i = O_{i+1}$. We have by induction that p_i is the unique element in locations $i+1,\dots,N$ whose index is less than $i+1$. This implies hypothesis (3). For the same reason, the depth property continues to be satisfied in T'_i. Also, $I_i = I_{i+1}$, because the inversions with respect to p_i were already counted in I_{i+1}, and the element occupying location i in T'_{i+1} (now T'_i, and considered in I_i for the first time) is either x or (by the depth property) has index less than i, so has no inversions with respect to any elements in locations $i+1,\dots,N$. Hypothesis (5) is unchanged.
- p_j is neither x nor p_i, and $p_i(T'_{i+1}) = p_i(T)$. (That is, p_i occupies location i in T and in T'_{i+1}.) In this case, T'_i will be obtained by interchanging p_i with p_j. We have from hypothesis (5)(a) that $p_i(S) > x(S)$, and (since $j < i$, but $p_j(T'_{i+1}) > p_i(T)$) $p_j(T'_{i+1}) \neq p_j(T)$, so $p_j(S) < x(S)$. Hence there is an interchange in O_{i+1} between p_i and p_j. Remove it to obtain $O_i \subset O_{i+1}$. The element p_j now occupies location i, element p_i now satisfies the depth property, and elements $p_i, p_{i+1},\dots,p_N$ (other than x) occupy locations $i+1,\dots,N$ (though not necessarily in that order), so hypothesis (3) is also satisfied. Furthermore, the inversions I_i are the same as in I_{i+1}, since p_j and p_i are the two elements with smallest indices among those occupying locations $i, i+1,\dots,N$ (so that swapping p_i for p_j replaced inversions involving p_j with inversions involving p_i, but introduced no new inversions), and p_j is now in location i (so that there are no inversions involving p_j). Finally, we note that, even though the location of p_j has changed as a result of the swap, hypothesis (5)(a) remains satisfied because p_j now occupies location $i \neq j$, $p_j(T) \neq p_j(T'_i)$ as before; and hypothesis (5)(b) does not apply to location i.
- p_j is neither x nor p_i, and $p_i(T'_{i+1}) \neq p_i(T)$. In this case, $p_i(T'_{i+1}) < p_i(T)$ (that is, p_i occupies a position closer to the front of the list T'_{i+1} than position i). We swap

p_i with the element (p_k, say) occupying position i in T'_{i+1}. We have from hypothesis (5)(b) that $p_k(S) < p_i(S)$, and here that $p_k(T'_{i+1}) = p_i(T) > p_i(T'_{i+1})$, so there is a swap between them. Remove it from O_{i+1} to obtain O_i. The depth property continues to be satisfied, as is hypothesis (3). The only element in locations $i+1,\ldots,N$ in T'_i that is not in locations $i+1,\ldots,N$ in T (that is, does not have index $\geqslant i+1$) is p_j. We have therefore introduced only one additional inversion (that between p_i and p_j) by reason of the progression from I_{i+1} to I_i. That additional inversion is offset by the swap between p_i and p_k that we have removed from O_{i+1} to obtain O_i. (This is the only case in which this construction requires this offset.) Thus hypothesis (2) remains valid. Finally, we show that hypothesis (5) remains valid for p_k, the element swapped with p_i. Suppose p_i occupied location l in T'_{i+1}. Then $p_i(S) > p_l(S)$ by hypothesis (5)(b) (induction), and $p_k(S) > p_i(S)$ by hypothesis (5)(b) (induction), so $p_k(S) > p_l(S)$, establishing (5)(b), and in particular $k \neq l$, establishing (5)(a).

This exhausts the possible cases for Proposition 11, and concludes the proof. □

References

[1] S. Albers, Improved randomized on-line algorithms for the list update problem, SIAM J. Comput. 27 (1998) 682–693.

[2] S. Albers, Private communication.

[3] S. Albers, B. von Stengel, R. Werchner, A combined BIT and TIMESTAMP algorithm for the list update problem, Inform. Process. Lett. 56 (1995) 135–139.

[4] S. Albers, J. Westbrook, Self-organizing data structures, in: Online Algorithms: The State of the Art, Fiat-Woeginger, Springer, Berlin, 1998.

[5] J.L. Bentley, C. McGeoch, Amortized analysis of self-organizing sequential search heuristics, Commun. ACM 28 (4) (1985) 404–411.

[6] A. Borodin, N. Linial, M. Saks, An optimal online algorithm for metrical task systems, J. ACM 52 (1995) 46–52.

[7] W.R. Burley, Traversing layered graphs using the work function algorithm, J. Algorithms 20 (3) (1996) 479–511.

[8] W. Burley, S. Irani, On algorithm design for metrical task systems, ACM-SIAM Symp. on Discrete Algorithms, 1995.

[9] M. Chrobak, L. Larmore, The server problem and on-line games, On-Line Algorithms, Proc. DIMACS Workshop, DIMACS Series in Discrete Mathematics and Computer Science, Vol. 7, 1991, pp. 11–64.

[10] M. Chrobak, J. Noga, Competitive algorithms for multilevel caching and relaxed list update, Proc. of ACM–SIAM Symp. on Discrete Algorithms, 1998.

[11] R. El-Yaniv, There are infinitely many competitive-optimal online list accessing algorithms, Discussion paper from The Center for Rationality and Interactive Decision Making, Hebrew University.

[12] S. Irani, Two results on the list update problem, Inform. Process. Lett. 38 (6) (1991) 301–306.

[13] E. Koutsoupias, C. Papadimitriou, On the k-server conjecture, J. ACM 42 (5) (1995) 971–983.

[14] M. Manasse, L. McGeoch, D.D. Sleator, Competitive algorithms for server problems, J. Algorithms 11 (1990) 208–230.

[15] N. Reingold, J. Westbrook, Off-line algorithms for the list update problem, Informat. Process. Lett. 60 (2) (1996) 75–80.

[16] S. Roura, C. Martinez, On the competitiveness of the Move-to-front rule, Technical Report LSI-96-63-R, Technical University of Catalonia, 1997.

[17] F. Schulz, Two new families of list update algorithms, Proc. Ninth Int. Symp., ISAAC, Lecture Notes in Computer Science, Vol. 1533, Springer, Berlin, 1998, pp. 99–108.

[18] D.D. Sleator, R.E. Tarjan, Amortized efficiency of list update and paging rules, Commun. ACM 28 (1985) 202–208.
[19] D.D. Sleatorm, R.E. Tarjan, Self-adjusting binary search trees, J. ACM 32 (1985) 652–686.
[20] B. Teia, A lower bound for randomized list update algorithms, Inform. Process. Lett. 47 (1993) 5–9.